Extension of the QUASAR river water quality model to incorporate dead-zone mixing

M. J. Lees, L. Camacho, P. Whitehead

Hydrology and Earth System Sciences, 2(2–3), *353–365* (1998) © EGS

Hydrology & Earth System Sciences

Extension of the QUASAR river water quality model to incorporate dead-zone mixing

Matthew J. Lees*, Luis Camacho* and Paul Whitehead†

*Department of Civil and Environmental Engineering, Imperial College of Science, Technology and Medicine, London, SW7 2BU, UK.
†Department of Geography, University of Reading, Reading, RG6 6AB, UK.

Abstract

A modification to the well-known water quality model 'Quality Simulation Along River Systems' (QUASAR) is presented, extending its utility to real-time forecasting applications such as the management and control of pollution incidents. Two aggregated dead-zone (ADZ) parameters, namely time delay and dispersive fraction, are incorporated into the existing model formulation, extending the current continuously stirred tank reactor based model processes to account for advective and active mixing volume dispersive processes. The resulting river water quality model combines the strengths of the QUASAR model, which has proven non-conservative pollutant modelling capabilities, with the accurate advection and dispersion characterisation of the ADZ model. A discrete-time mathematical representation of the governing equations is developed that enables efficient system identification methods of parameter estimation to be utilised. The enhanced water quality model and associated methods of parameter estimation are validated using data from tracer experiments conducted on the River Mimram. The revised model produces accurate predictions of observed concentration-time curves for conservative substances.

Introduction

QUASAR (Whitehead *et al.*, 1997) provides a quantitative framework to analyse the physical, chemical and biological aspects of complex river systems. In its present state the model provides a means of assessing current environmental impacts of pollutants on water quality (Whitehead and Williams, 1982; Whitehead *et al.*, 1984); of carrying out general research on those aquatic systems (Ferrier *et al.*, 1995); of addressing future input scenarios (Whitehead *et al.*, 1995); and assisting in the development of practical management plans of river systems (Lewis *et al.*, 1997; Whitehead *et al.*, 1979 and Whitehead *et al.*, 1981) including the setting of quality levels of discharges to meet river quality objectives. A typical application, in which the model may be suitable, is the analysis of impacts upon water quality resulting from changes in flow regime.

Despite inherent difficulties that arise when modelling large river systems, it has been possible to calibrate and validate the model with long term observed data, and it has been shown that the major processes affecting DO-BOD interactions, as well as those affecting nitrate (e.g., nitrification and denitrification) are correctly formulated (see Lewis *et al.* 1997). However, one existing problem with the current model is the use of continuously stirred tank reactors (CSTR) in series to represent the pollutant advection and dispersion processes. This model formulation, sometimes referred to as a cells in series model (Rutherford, 1994), is not the most appropriate for applications that require short term forecasts, such as in the assessment and real-time prediction of pollution events, because an unknown number of serial connections is required to represent observed advection and dispersion in a single river reach.

This paper presents an extension to QUASAR that introduces aggregated dead-zone (ADZ) concepts (Beer and Young 1983; Wallis *et al.*, 1989a,b; Young and Wallis 1993), in order to overcome this problem. Extension of the model requires modification of the CSTR approach to incorporate an advective term to account for observed advective time delay and a term to characterise the fraction of the reach volume that can be considered fully mixed, the active mixing volume (AMV, see Young and Lees, 1993). In ADZ terminology the volumetric fraction of the aggregated dead-zone or AMV to the total reach volume is termed the dispersive fraction.

The paper presents a methodology for the estimation of the advection and dispersion parameters using appropriate modifications of two approaches developed for the ADZ model. The first approach is a simple subjective method which uses derived relationships from observed concentration-time data from two downstream locations (Wallis *et al.*, 1989b); while the second objective method is based on

the Simplified Refined Instrumental Variable (SRIV, see, Young 1984, 1992) method of system identification. The latter method is superior since it allows the modeller to identify an appropriate model structure in addition to estimating the coefficients of a discrete transfer function model, from which the dispersive fraction and the time delay can be accurately computed.

The paper is structured as follows. The existing QUASAR model formulation of advection, dispersion and decay processes is presented and compared with the ADZ model. Next, dead-zone extensions to the model are presented along with suitable methods of parameter estimation for the newly incorporated dead-zone parameters. Finally, the enhanced model, which is implemented using the continuous simulation SIMULINK software package (Mathworks, 1996), is tested using data from tracer experiments conducted in the River Mimram.

Modelling advection, dispersion and decay processes in streams

In order to provide the relevant background to the extended QUASAR model, brief descriptions of the QUASAR and ADZ models are given below.

THE QUASAR MODEL

The water quality model QUASAR is a simple dynamic model for non-tidal rivers that comprises a set of ordinary differential equations describing the changes of flow and concentration of different water quality determinands over time. The reader is directed to Whitehead *et al.* (1997) for a detailed description of QUASAR's theory and development, since only relevant aspects are discussed here. A simple multi-reach approach is used to model river systems where each river reach is considered as a set of completely mixed systems, or CSTRs in series, in which chemical decay processes and biological behaviour are simulated. For each reach, time-varying flow inputs from tributaries, point and diffuse effluent discharges and pollutant inputs as well as flow abstractions, can be taken into account. Figure 1 shows a QUASAR conceptualisation of such a river section.

The specific water quality equation for each determinand results from the application of the conservation of mass principle over a completely mixed reach (see Whitehead *et al.*, 1997). Under steady-state flow conditions and assuming that the loss of solute by reaction or settling can be represented by a simple proportional mass decay, the mass balance takes the form:

$$S\frac{dX(t)}{dt} = Q \cdot U(t) - Q \cdot X(t) - k \cdot S \cdot X(t) \qquad (1)$$

where $U(t)$ is the input concentration of the water quality determinand; $X(t)$ is the output concentration; Q is the flow; k is the decay or sedimentation rate coefficient; and S is the volume storage in the reach, which can be related to flow using the following storage equation,

$$S = \bar{t} \cdot Q \qquad (2)$$

Here $\bar{t}$ is the travel time parameter that is commonly calculated (see Chapra, 1997) from mean velocity estimates or measurements $\bar{v}$ and the longitudinal reach length L as,

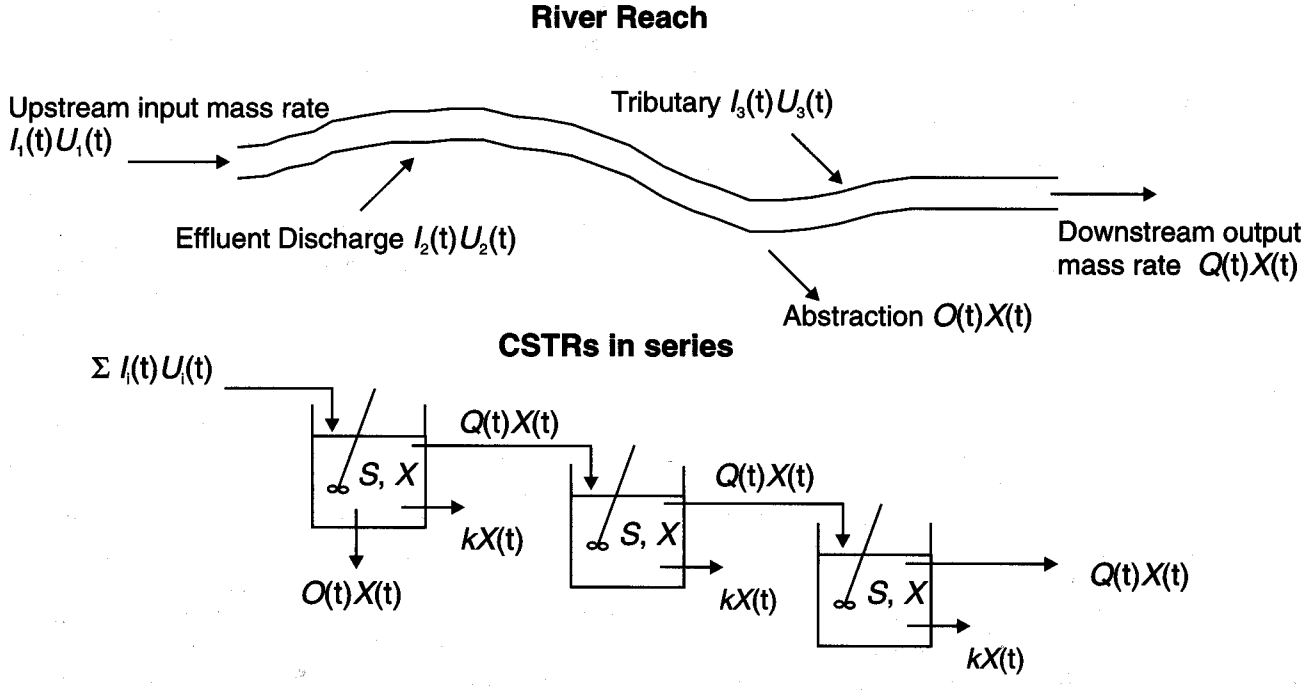

Fig. 1 *QUASAR conceptualisation of a river reach*

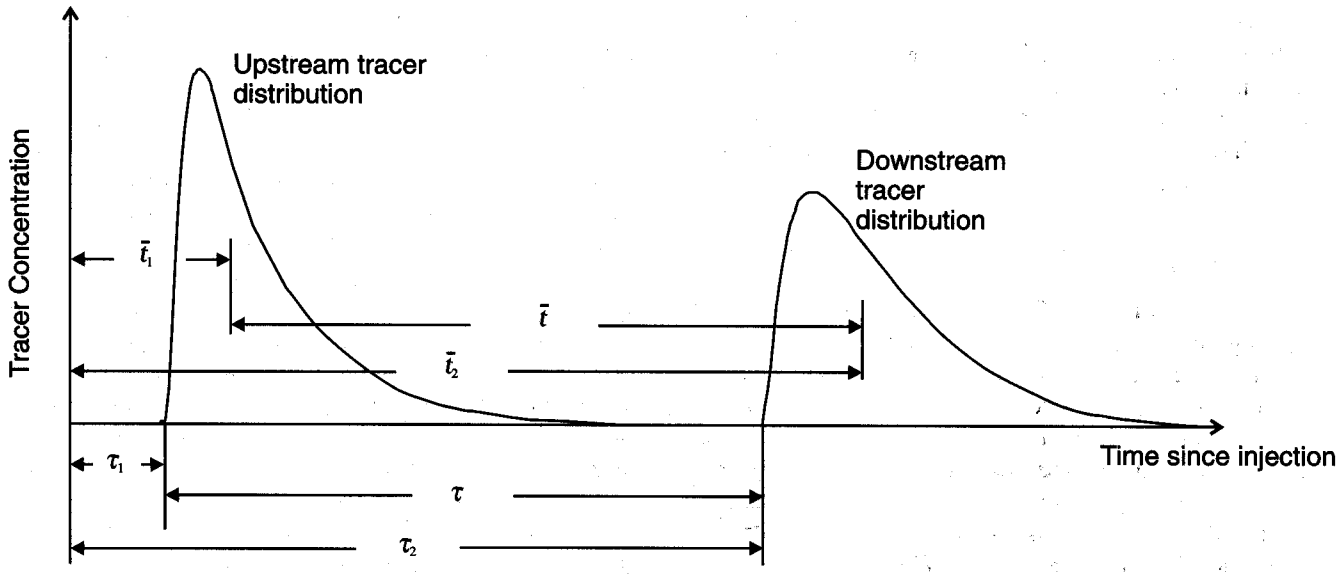

Fig. 2 *Diagrammatic representation of travel time $\bar{t}$ and time delay τ*

$$\bar{t} = L / \bar{v} \tag{3}$$

Note that in steady state flow conditions the velocity is constant over time and thus Eqn. 2 represents a linear relationship. Either point estimates of mean velocity at various cross-sections or reach estimates from tracer experiments may be used.

For a general water quality determinand Eqn. 1 simplifies to the general form,

$$\frac{dX(t)}{dt} = \frac{1}{\bar{t}}(U(t) - X(t)) + \sum sources - \sum sinks \tag{4}$$

in which the sources and sinks can represent transport mechanisms, e.g. settling, volatilisation, etc., but mainly correspond to biochemical decay processes. The water quality determinands considered in QUASAR include nitrate, dissolved oxygen (DO), biochemical oxygen demand (BOD), ammonium ion, temperature, pH, and conservative substances.

THE AGGREGATED DEAD-ZONE MODEL

Beer and Young (1983) introduced the concept of an aggregated dead-zone (ADZ) model to represent the dispersive effects of all 'dead' or 'storage' zones in a river reach. All dispersive processes that a solute undergoes in a river reach are lumped in a single parameter, the aggregated dead-zone residence time T_r. The assumption underlying the ADZ model is that the major dispersive effects reside in the effective dead-zone residence-time and not in the Fickian diffusion term of the advection-diffusion equation (ADE) proposed by Taylor (1954). Therefore, a single ordinary differential equation representation, in which the coefficient of dispersion no longer appears, is used to describe the overall dispersive effects. A comprehensive description of the ADZ model resulting from these initial concepts can be found in Wallis *et al.* (1989a) and Young and Wallis (1993).

The mass balance equation for a general solute in a fully mixed aggregated dead-zone or AMV with volume V takes the form,

$$V\frac{dX(t)}{dt} = Q \cdot U(t) - Q \cdot X(t) - k \cdot V \cdot X(t) \tag{5}$$

where the variables are as defined previously. It is worth noting the similarity between Eqns. 1 and 5. Both equations are essentially the same except that the mass balance is performed for different fully mixed control volumes, i.e. V is the active mixing volume and S is the total reach volume. In fact, the dispersive fraction,

$$DF = \frac{V}{S} \tag{6}$$

is the ADZ parameter that defines dispersive effects in the stream, and can be interpreted as a measure of the fractional volume of the reach responsible for dispersion. As in QUASAR, a linear storage relationship is assumed to relate the AMV storage volume to the hydro-geometric flow parameters,

$$V = T_r \cdot Q \tag{7}$$

where T_r is defined as the residence time of the solute in the aggregate dead-zone. From Eqns. 2 and 7 it can be seen that the dispersive fraction can also be described by,

$$DF = \frac{T_r}{\bar{t}} \tag{8}$$

revealing that this parameter can also be interpreted as the time that the tracer spends being dispersed in the reach as a fraction of the total time it spends in the reach.

In addition to the process of dispersion, Beer and Young (1983) proposed the introduction of a pure time delay of τ time units to model solute advection between the upstream and downstream sites due to bulk flow movement. Introducing this term and considering Eqn. 7, the ADZ model takes the form,

$$\frac{dX(t)}{dt} = \frac{1}{T_r}[U(t-\tau) - X(t)] - k \cdot X(t) \qquad (9)$$

where, as shown in Fig. 2, τ is the time taken for the leading edge of the solute cloud to be advected through the reach. However, the model is not yet complete since, for the non-conservative situation under consideration, it is necessary to include the mass decay of the solute occurring during the time period associated with the advection process. Adopting the procedure proposed by Young and Wallis (1993), the processes of advection, dispersion, and proportional decay for a general pollutant in a river reach can be described by the following equation,

$$\frac{dX(t)}{dt} = \frac{1}{T_r}[e^{-k\tau} \cdot U(t-\tau) - X(t)] - k \cdot X(t) \qquad (10)$$

Comparison of Eqns. 4 and 10 reveals that the models are very similar with the exception that the QUASAR model assumes there is no pure advection and that the reach is fully mixed, whereas the ADZ model includes an explicit advection parameter and only assumes a proportion of the reach to be fully mixed. For many natural rivers the ADZ model has been shown to reproduce observed tracer responses accurately (e.g. Wallis; 1994, Rutherford 1994) supporting the conceptual formulation of the model. To ensure correct characterisation of observed pollutant transport, the QUASAR reach model can be formulated as a number of sub-reach models in series (Himelblau and Yates, 1968; Weinmann and Laurenson, 1979; Whitehead *et al.*, 1979). However, since there is no explicit time delay parameter to account for observed advection between upstream and downstream sampling sites, use of the model for forecasting and real time modelling applications is problematic. The number of cells or sub-reaches into which each reach is sub-divided must be determined by calibration to ensure that the patterns of advection and dispersion are correctly simulated. In other words, the time delay is implicitly taken into account in the geometric configuration of the river system adopted. Different configurations, however, result in a wide variety of *non-intuitive* dispersive and advective patterns.

As an illustrative example of this limitation, Fig. 3 shows the results of a conservative pollutant simulation using various numbers of sub-reaches to simulate the advective and dispersive processes in a river reach 3600 meters long. A two hour long pulse of 200 mgl^{-1} is injected at the upstream end, and the downstream concentration curve is computed using various configurations of sub-reaches. It is evident from Fig. 3 that the number of sub-reaches used in a simulation has a direct effect on the reproduction of advection and dispersion processes. The configuration, therefore, must be investigated carefully for each river reach and flow condition resulting in increased calibration costs. In this context an explicit parameter to control the observed time delay, thereby modelling the advection process appropriately, is advantageous.

In the case of the ADZ model, it should be noted that in the situation where the point source pollutant is not well

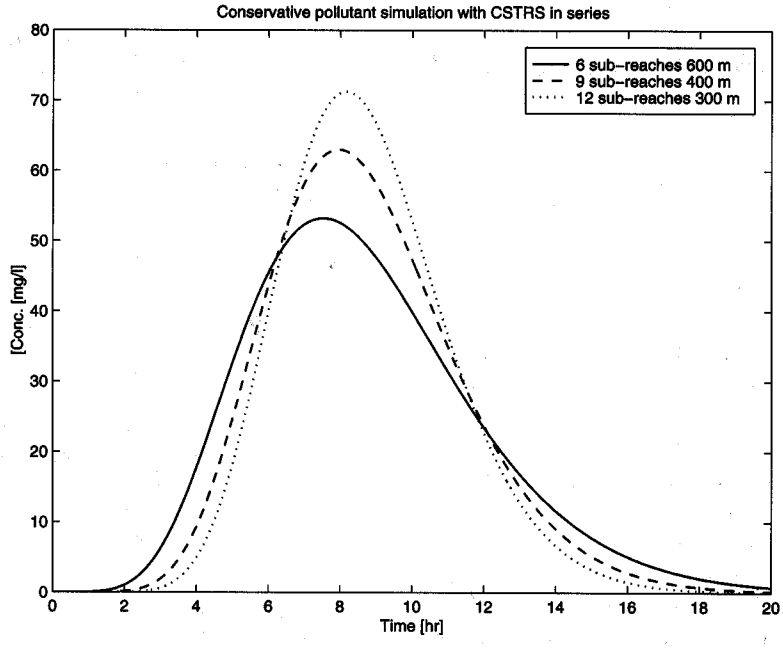

Fig. 3 *Simulation results of the QUASAR model with a single 3600 m reach divided into different numbers of sub-reaches*

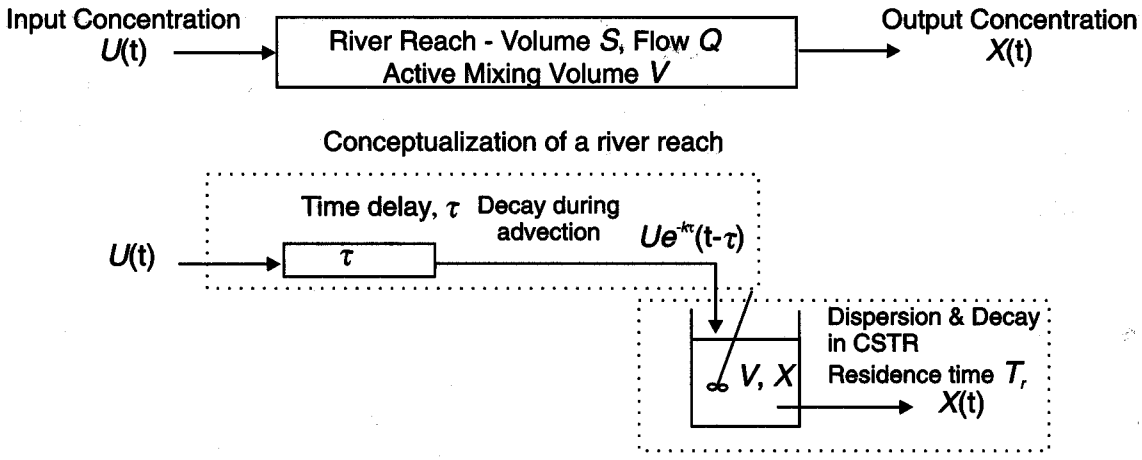

Fig. 4 *Conceptualisation of river reach transport mechanisms*

mixed laterally, a number of reach elements described by Eqn. 10 maybe required. The number of such elements will depend on the length and dispersive characteristics of the system being modelled which again are not known *a priori*. However, as discussed later, this multi-reach characterisation does not cause problems, since objective system identification techniques can be used to identify the required structure. In particular, lateral mixing effects can have a large influence on the characteristics of the downstream skewed curve for impulse inputs. This means that first order ADZ model results in such cases may differ substantially from the observed concentration-time curves. To overcome this practical problem Green and Beven (1993) suggest that data for ADZ modelling analysis should ideally consist of measurements of concentration-time curves at two or more downstream sites, defining input and output curves for a reach.

Extension of QUASAR to include dead-zone concepts

In the previous two sections the QUASAR and ADZ model structures have been presented, revealing similarities in their basic model formulations (cf. Eqn. 4 and Eqn. 10). This section presents an extension of the QUASAR model to incorporate dead-zone mixing processes as conceptualised in the ADZ model.

Figure 4, which is a diagrammatic representation of a first order ADZ reach, shows that advection and dispersion processes of a solute cloud in a channel are treated independently and can be characterised by two ADZ model parameters, T_r and τ. As stated previously the advective time delay τ is the time taken for the leading edge of the pollutant cloud to be advected through the reach, and the mixing residence time parameter T_r is the mixing time associated with the solute passage through the aggregated dead-zone volume. Note that the decay process occurs during advection and dispersion resulting in decay taking place over an equivalent time period to that considered in QUASAR since $\bar{t} = T_r + \tau$.

Introducing the ADZ parameters T_r and τ into the QUASAR model gives,

$$\frac{dX(t)}{dt} = \frac{1}{T_r}\left(e^{-k\tau} \cdot U(t-\tau) - X(t)\right) = \sum sources - \sum sinks \tag{11}$$

where the pure time delay can be calculated by,

$$\tau = \bar{t} \cdot (1 - DF) \tag{12}$$

$$\bar{t} = \frac{L}{\bar{v}} \tag{13}$$

and the residence time of the aggregated dead-zone or active mixing volume from Eqn. 7 and Eqn. 12 is,

$$T_r = \bar{t} - \tau = \bar{t} \cdot DF \tag{14}$$

As an example, consider the modified QUASAR equations describing the behaviour of a conservative determinand X_c and the non-conservative pollutant ammonium X_{amm}. These now read, respectively,

$$\frac{dX_c(t)}{dt} = \frac{U_c(t-\tau) - X_c(t)}{T_r} \tag{15}$$

$$\frac{dX_{amm}(t)}{dt} = \frac{e^{-K_1\tau}U_{amm}(t-\tau) - X_{amm}(t)}{T_r} - [K_1 X_{amm}(t)] \tag{16}$$

where: U_c is the input concentration of the conservative water quality determinand
X_c is the output concentration
T_r is the ADZ residence time
U_{amm} is the input ammonium concentration
X_{amm} is the output ammonium concentration
K_1 is the nitrification rate at a given temperature, $K_1 = K_1^0 \cdot 10^{(0.02937 T_0)}$
K_1^0 is the nitrification rate coefficient at 20 °C
T_0 is temperature

The extended QUASAR model can be simulated using

any suitable numerical integration algorithm such as the Runge-Kutta 4th order method. Here we have coded the model using the iconographic simulation package SIMULINK (Mathworks, 1996). The resulting iconographic simulation model can be easily connected together to represent complex river systems without the requirement of a traditionally programmed graphical user interface (Camacho, 1997).

Estimation of the aggregated dead zone parameters

Two methods, based on the results of planned tracer experiments, can be used to estimate the new aggregated dead zone parameters that have been incorporated into the model. The first subjective method considers simple relationships obtained from the observed concentration-time data at two locations downstream of the tracer injection point, while the second method, which is based on time series analysis techniques, provides objective model identification and parameter estimation.

Wallis *et al.* (1989a,b) investigated the link between the ADZ model parameters and conventional open channel flow parameters. In the case where a simple reach model explains observed dispersion patterns, T_r can be calculated from a pair of observed concentration-time curves (see Fig. 2) using,

$$T_r = \bar{t} - \tau \tag{17}$$

where $\bar{t}$ is the mean travel time along the reach, computed as the difference between the times of passage of the dye centroids past the downstream and upstream sites ($\bar{t} = \bar{t}_2 - \bar{t}_1$); and τ is the time delay computed as the difference between the first arrival times of tracer at the same sites ($\tau = \tau_2 - \tau_1$). In the case of a conceptualised serial sub-reach system, T_r is equal to the sum of the individual element residence times, and in the case of a parallel sub-reach system it is equal to a weighted average of the individual residence times (Young, 1992; Young and Wallis, 1993).

The second approach to parameter estimation involves a reformulation of the differential equations as difference equations (Wallis *et al.*, 1989a) enabling application of flexible methods of time-series analysis for model identification and estimation. Following the comprehensive presentations by Young (1992) and Young and Wallis (1993), a general discrete-time transfer function model, which is applicable to a number of ADZ elements connected in series or parallel in a river system, can be represented as:

$$X_k = \frac{B(z^{-1}) \cdot z^{-\delta}}{A(z^{-1})} U_k + \xi_k \tag{18}$$

where k denotes the kth sampling instant; U_k is the input concentration at the kth sample; X_k is the output concentration; ξ_k is a random variable introduced to allow for unmodelled stochastic effects such as measurement errors; δ is the advective time delay in sampling intervals or time steps Δt, defined as the nearest lower integer value of $\tau/\Delta t$; and $B(z^{-1})$ and $A(z^{-1})$ are backward shift operator polynomials,

$$B(z^{-1}) = b_0 + b_1 \cdot z^{-1} + \ldots + b_m \cdot z^{-m} \tag{19}$$

$$A(z^{-1}) = 1 + a_1 \cdot z^{-1} + \ldots + a_n \cdot z^{-n} \tag{20}$$

where the backward shift operator is defined as z^{-p}, $z^{-p}x_k = x_{k-p}$.

The problem of model structure identification involves determining the numbers of a and b parameters (n, $m+1$) and δ, denoted as the triad [n, $m+1$, δ] ($n = 1,2,\ldots$; $m = 0,1,2,\ldots$; $\delta = 0,1,2,\ldots$) that best describe the relationship between the observed time-series U_k and X_k at upstream and downstream locations. Correspondingly, once a model order is identified the problem of estimation involves the determination of the $a_1 \ldots a_n$ and $b_0 \ldots b_m$ parameters.

Transfer function model parameters can be estimated using the Simplified Refined Instrumental Variable (SRIV) method of parameter estimation (Young, 1984, 1991). Model identification is performed using a procedure based on the repeated evaluation of different statistical measures for different model orders, and the selection of the 'best' identified model as that producing the best criteria. The statistics recommended by Young (1992) for a successful identification procedure are Young's information criterion (*YIC*) and the coefficient of determination R^2. The *YIC* gives a measure of model fit and parametric efficiency, whereas R^2, which is the percentage of the output data variance explained by the model, is solely a fit criterion. If the model explains the data well and is not over-parameterised, the value of R^2 approaches unity and the *YIC* becomes largely negative.

As indicated by Young (1992) if the denominator polynomial $A(z^{-1})$ has real eigenvalues then it can always be interpreted, unambiguously, as a series and/or parallel connection of first and second-order subsystems. For example, a second order model [2,1,δ] can be factorised into two first order models [1,1,$p\delta$] and [1,1,$(1-p)\delta$] representing two reaches in series. Here p is a fraction of the time delay in sampling intervals, *i.e.* $0<p<1$, that takes the value 0.5 for reaches of the same length and hydraulic characteristics. In a similar manner, a second order model of the form [2,2,δ] can represent a parallel connection of two first order systems.

Once the parameters of a first order model have been estimated, the residence time and the dispersive fraction can be calculated. Note that the single 'a' parameter of the first order transfer function model can be approximately related to the residence time of the continuous-time, conservative model by,

$$T_r = -\frac{\Delta t}{\log_e(-a_1)} \tag{21}$$

In the non-conservative case,

$$k + \frac{1}{T_r} = -\frac{\log_e(-a_1)}{\Delta t} \tag{22}$$

The mean travel time can then be computed as,

$$\bar{t} = \Delta t\delta + T_r \tag{23}$$

where $\Delta t\delta$ is the identified time delay. Thus, the dispersive fraction is given by $DF = T_r/\bar{t}$. In addition, it is also possible to estimate the steady-state gain (SSG) of the first order model as,

$$SSG = \frac{b_0}{1 + a_1} \tag{24}$$

The value of the SSG can be used to check if the tracer behaves in a conservative manner. Values of SSG less than 1 indicate either decay of the tracer between sampling stations; or that there are tributary inputs or additional dilution sources not accounted for; or that complete lateral mixing has not been attained at the measurement sites.

As stated earlier, both methods of parameter estimation produce better results if the tracer cloud is measured at two points downstream of the injection point. This is because, in this case, often only a single ADZ model element is required to describe the observed input-output relationship therefore simplifying the physical interpretation of the modelling results. Higher order models are often identified where an impulse input is used because of the initial high dimension mixing effects (Green and Beven, 1993).

One advantage of the SRIV parameter estimation method is that it not only produces unbiased estimates of the transfer function parameters (a_1, a_2, ... a_n, b_0, b_1, ... b_m), but it also provides an estimate of parameter uncertainty in the form of standard errors, which can be easily calculated from the error covariance matrix P^* (Young, 1984). To relate the uncertainty of these transfer function parameters to the mechanistic model parameters (Eqn. 10), a Monte Carlo (MC) simulation technique can be utilised (Young, 1992). In this MC simulation the transfer function model parameters are selected randomly from a probability distribution defined by the SRIV estimation results, *i.e.* with mean and variance defined by the estimated coefficients and error covariance matrix. The continuous-time model parameters T_r, $\bar{t}$ and DF are then calculated for each random realisation (Eqn. 21, Eqn. 23 and Eqn. 8) producing probability distributions for each parameter.

Model validation

Since the main motivation behind the ADZ extension of QUASAR is to improve the characterisation of advection and dispersion processes, it is appropriate to test the model using data collected from planned tracer experiments. Although this validation is limited to conservative pollutants, it validates the representation of transport mechanisms within the extended model, which we believe, should be correctly characterised prior to consideration of the chemical and biological decay processes that affect non-conservative pollutants. To this end, data from a conservative tracer experiment carried out on the River Mimram are used to validate the proposed extended model.

RIVER MIMRAM TRACER EXPERIMENTS

A tracer experiment was carried out on the River Mimram near the Panshanger flow gauging flume in Hertfordshire, England. The reach is approximately 200 meters long and is characterised by non-uniform cross-sections of sandy-pebbled bed with heavy weed growth. Approximately 10 kg of Sodium Chloride was gulp injected into the river upstream of the flume's hydraulic jump, and the resulting tracer cloud was measured over time at three sampling stations downstream as follows: site A, 100 metres downstream from the injection point; site B, 40 metres downstream from site A; site C, 50 metres downstream from site B. Measurements of conductivity were taken at irregular time intervals at each cross-section, and the concentration of Sodium Chloride was computed from the conductivity meter calibration curves. The recorded concentrations, interpolated over a uniform sampling interval of 10 seconds, are shown in Fig. 5.

SRIV MODEL IDENTIFICATION AND ESTIMATION

The data from sampling stations A and B define the upstream input concentration for reaches B and C respectively, while a unit impulse input is assumed as the input to reach A. Table 1 summarises the best results obtained from the SRIV identification and estimation of model structures of different orders ranging from [1,1,1] to [3,3,20] for both reaches.

The results suggest that first order models describe the observed outputs exceptionally well. For Reach B, the [1,1,9] model is well identified with a largely negative YIC and it explains over 99% of the output data variance. In this case, the model steady gain (SSG) is given by (see Eqn. 24),

$$SSG = 0.1995/(1 - 0.7905) = 0.952$$

indicating a 95% mass conservation between sampling stations A and B. The ADZ residence time (T_r) is then given by (see Eqn. 21),

$$T_r = \frac{-\Delta t}{\log_e(-a_1)} = \frac{-10}{\log_e(0.7905)} = 42.54s \tag{a}$$

The time delay and the mean travel time can then be calculated,

$$\tau = \Delta t \cdot \delta = 10 \cdot 9 = 90s; \quad \bar{t} = 90 + 42.54 = 132.54s \tag{b}$$

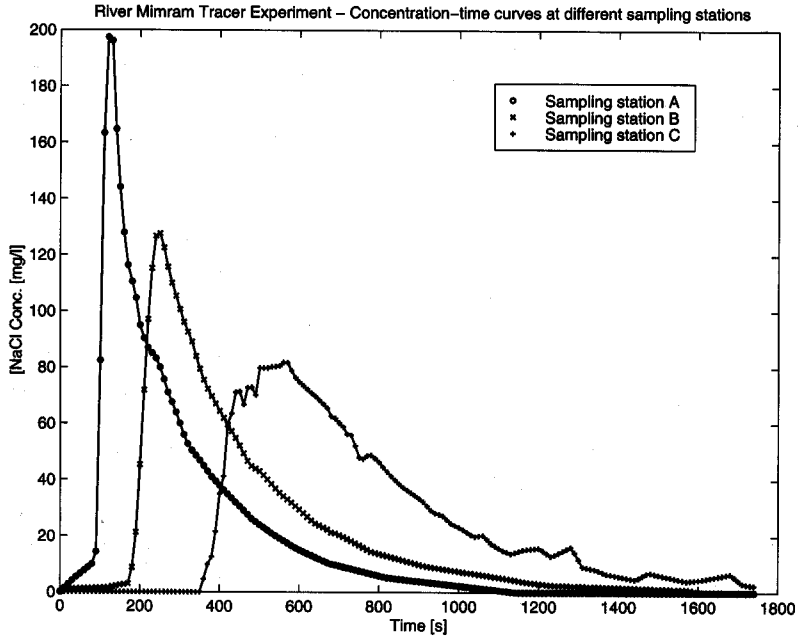

Fig. 5 *Observed concentration-time curves at stations A, B, and C*

The dispersive fraction results in,

$$DF = 42.54/132.54 = 0321$$

In addition, for this 40 m reach the mean reach velocity is,

$$\bar{v} = \frac{L}{\bar{t}} = \frac{40}{132.54} = 0.302\text{ms}^{-1} \qquad \text{(c)}$$

The estimated aggregated dead-zone residence time and the advective pure time delay can then be inserted into the extended QUASAR simulation model allowing a predicted downstream concentration to be calculated. Figure 6, which shows the extended QUASAR model result, reveals that the model is explaining the dispersion and advection properties of this reach exceptionally well. In particular note how the recession part of the curve, and the overall skewed shape of the profile, are well reproduced by the model. Figure 6 also shows the model simulation results for Reach C. Again, it can be seen that the model output is very close to the observed tracer curve.

For comparison, the dead zone model parameters have also been estimated using the subjective method suggested by Wallis *et al.* (1989b). The results prove to be acceptable, although not as good as the approach used above. However, in situations where no access to objective time series analysis estimation is possible, this method could be

Table 1. SRIV identification and estimation results for reaches B and C. (Standard errors in parentheses)

$[n, m+1, \delta]$	*YIC*	R^2	*a*	*b*
REACH B				
1, 1, 9	–13.42	0.994	–0.791(0.0035)	0.20(0.0031)
2, 2, 10	–13.09	0.996	–1.652(0.0084), 0.660(0.0075)	0.276(0.0041), –0.268(0.0038)
3, 1, 7	–11.20	0.994	–2.333(0.022), 1.937(0.037), –0.574(0.0163)	0.029(0.001)
REACH C				
1, 1, 16	–13.88	0.993	0.925(0.001)	0.085(0.001)
2, 1, 17	–13.46	0.982	–0.533(0.001), –0.376(0.001)	0.105(0.0093)
3, 1, 17	–15.35	0.994	–1.810(0.001), 1.430(0.002), –0.559(0.001)	0.064(0.0003)

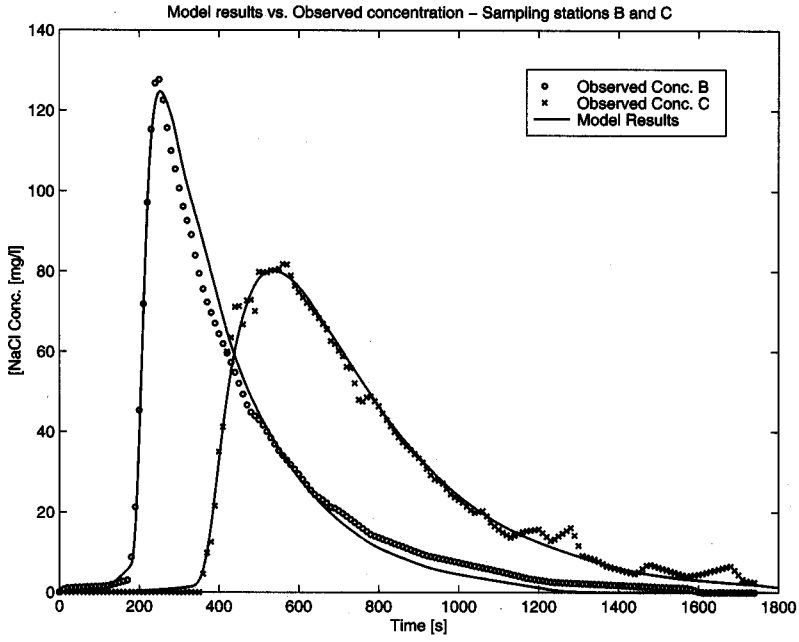

Fig. 6 *Extended QUASAR results with dead-zone parameters estimated using the SRIV method*

used with some confidence. This is particularly true when data are available at more than one sampling station, and a first order model is able to explain the downstream observed curves.

MODEL COMPARISON

To illustrate how the proposed model compares with the existing QUASAR model, the same data were also modelled using the conventional CSTR representation. For the case of the conservative solute under consideration, the conventional QUASAR model (Eqn. 4) simplifies to,

$$\frac{dX(t)}{dt} = \frac{1}{\bar{t}}\big(U(t) - X(t)\big) \qquad (25)$$

Here, calibration requires the identification of the travel time parameter T for each tank in a serial system of n tanks described by Eqn. 25 that best reproduce the observed downstream time series of data. For a fixed mean travel time $\bar{t}$ along a reach, the corresponding travel time parameter of each identical tank is given by $T = \bar{t}/n$. In the present example the mean travel times identified using SRIV system identification for Reaches B and C are $\bar{t}_B$ = 132.54s and $\bar{t}_C$ = 288.45s respectively. In previous practical QUASAR applications, this overall travel time parameter has been estimated either through a trial-and-error or numerical optimisation procedure. Application of an automatic optimisation procedure in this case confirmed that the values calculated using the SRIV method produce the best explanation of the observed tracer curves. Simulation results obtained for various serial configurations are summarised in Table 2, and the best identified results (n_B = 10, n_C = 6) are compared in Fig. 7 against the observed time series. Note that there is little to choose between different structures in terms of model fit, with no clear best model structure. The reason for this lack of uniqueness can be seen in Fig. 8 which shows the model outputs for a number of different CSTR configurations. This figure clearly demonstrates the difficulty of simultaneously reproducing the observed time delay, the peak and the recession curves of the highly skewed concentration-time distributions using the conventional QUASAR model.

This observation is not new and has already being pointed out by Rutherford (1994) for the cells in series contaminant transport model and by Doodge (1973) for the case of the Nash cascade flow routing model. The explanation comes from the fact that the number of tanks determines both the travel time, the rate of dispersion and the

Table 2. QUASAR-CSTR system identification results

Number of CSTRs [n]	Travel time $\bar{t}/n$ [s]	R^2
REACH B		
9	14.730	0.9670
10	13.254	0.9673
11	12.049	0.9670
15	8.836	0.9627
REACH C		
5	57.690	0.9573
6	48.075	0.9588
7	41.207	0.9560
8	36.056	0.9509

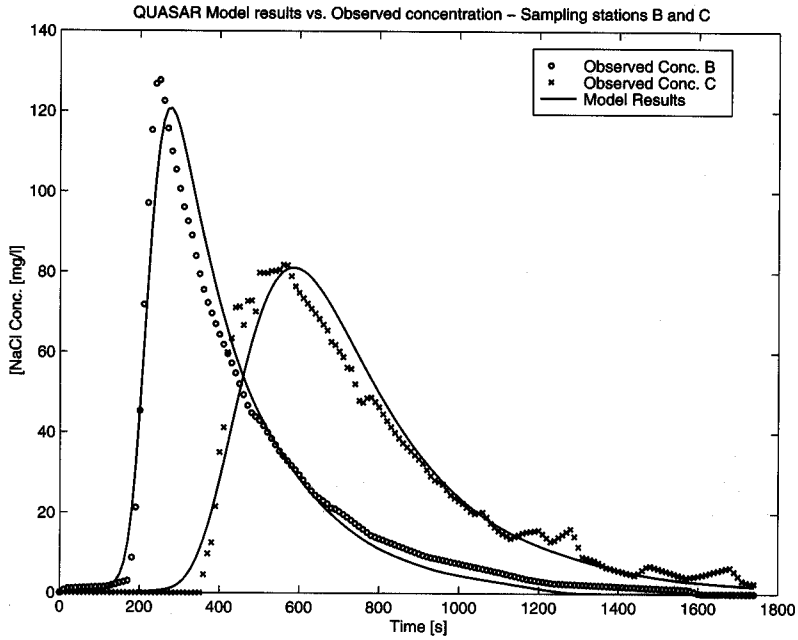

Fig. 7 *Existing QUASAR based model results for stations B and C*

skewness. The parameters cannot be varied independently as is evident from examination of the first three moments of the CSTR model,

$$\bar{t} = nT \quad \text{1st moment about the origin (Travel time)} \tag{26}$$

$$\sigma_t^2 = nT^2 \quad \text{2nd moment about centre of area (Variance)} \tag{27}$$

$$g_t = 2nT^3 \quad \text{3rd moment about centre of area (Skewness)} \tag{28}$$

In the case of the ADZ model, the additional time delay parameter provides the required flexibility since, although the last two moments remain unchanged in form, the first moment of a serial system (cf. Eqn. 17) is given by $\bar{t} = nT_r + \tau$. In addition, the assumption of incomplete mixing implicit in the residence time parameter T_r also contributes to the ADZ model's ability to describe skewed distributions, since this parameter is estimated objectively from observed data.

MODEL ESTIMATION IN A LIMITED DATA SITUATION

To illustrate further the power of the SRIV method of parameter estimation, the analysis is repeated under the assumption that data are only available at a single site downstream of the impulse injection point. This is a situation unfortunately very often encountered in practice. In this example, it is assumed that data are available only from the downstream end of Reach C. Table 3 summarises the SRIV identification and estimation results for a model structure identification range as before. The results show, as expected due to the impulsive input, that higher order models are now required to explain the downstream concentration curve adequately. In the present case, an excellent fit is obtained with a third order model [3,1,2] which is characterised by real eigenvalues (see Table 3). This model can be factorized into 3 first order models in series as represented diagrammatically in the following transfer function form,

Table 3. SRIV identification and estimation results for the combined reach (Standard errors in parentheses)

$[n, m+1, \delta]$	YIC	R^2	Eigenvalues	a	b
3, 1, 33	–15.224	0.992	0.968, 0.901, 0.377	–2.246(0.0016), 1.58(0.0031), –0.329(0.0015)	8.249 (0.021)
2, 1, 33	–13.718	0.990	0.969,0.902	–1.870(0.0014), 0.873(0.0014)	12.90 (0.133)
1, 1, 38	–9.359	0.895	0.979	–0.979(0.0009)	91.71 (2.695)

Third order model

$$u_{k-33} \rightarrow \left[\frac{8.2491}{1-2.246z^{-1}+1.577z^{-2}-0.329z^{-3}}\right] \rightarrow x_k \quad \text{(d)}$$

Three first order models in series (e)

$$u_{k-33} \rightarrow \left[\frac{2.0205}{1-0.9678z^{-1}}\right] \rightarrow \left[\frac{2.0205}{1-0.3769z^{-1}}\right] \rightarrow \left[\frac{2.0205}{1-0.9011z^{-1}}\right] x_k$$

The residence time parameters for the three reaches are $T_1 = 305.53$s, $T_2 = 10.25$s, $T_3 = 96.03$s, and the total pure time delay τ for the third order reach is 330s. At this point, there are various alternatives to factorising the third order reach into 3 serially connected first order reaches. An infinite number of reaches can be derived by choosing different time delays, i.e. all summing to 330s, representing reaches of different lengths or different physical characteristics. Different options could arise depending on the availability of hydrogeometric information for the river section under study. Ideally, reaches with uniform cross-sections should be identified. In the present case, however, a detailed hydrogeometric investigation of the river section was not carried out. Assuming that the 3 first order reaches have the same length as the reaches between measurement sites, we can use the previously identified time delays for reaches B and C, i.e., $\tau_2 = 90$, $\tau_3 = 160$, can be used to calculate the time delay of the third reach as $\tau_1 = 330 - 90 - 160 = 80$s. With the residence time and time delay parameters now defined, the other variables required for simulation can now be calculated.

Table 4 summarises the resulting model parameters using this assumption. It is interesting to note the wide variation in dispersive fractions arising from this factorisation. This variation can be tentatively related to the physical characteristics of the reaches and initial mixing characteristics. However, as mentioned previously, no firm links between the dispersive fraction and hydrodynamic parameters have so far been derived. Figure 9 shows the model results at sampling station C for three extended QUASAR models in series using the parameters identified from the decomposition, with a 20s duration impulse as the upstream input. The simulation results are excellent and demonstrate the power of this data-based mechanistic modelling method (Young *et al.*, 1996). Moreover, with parameters estimated solely from data collected at site C, the results for the intermediate sampling stations A and B show a good agreement with the observed data with R^2s of 0.89 for reach A and 0.97 for reach B. The model's ability to reproduce these intermediate observations is an indication that the factorisation of a third order model, with some knowledge about the physical characteristics of the reach, can result in accurate first order models for prediction at lower longitudinal scales. Here, previous knowledge of the time delays for the different intermediate reaches was obviously the key to achieving good results.

Table 4. Derived parameters for three reaches in series

Parameter	Reach 1	Reach 2	Reach 3
Residence Time T [s]	305.53	10.25	96.03
Time delay τ [s]	80	90	160
Mean travel time $\bar{t}$ [s]	385.53	100.25	256.03
Dispersive Fraction DF	0.7925	0.1022	0.3751
Reach Length L [m]	100	40	50
Mean Velocity $\bar{v}$ [m/s]	0.2594	0.399	0.1953

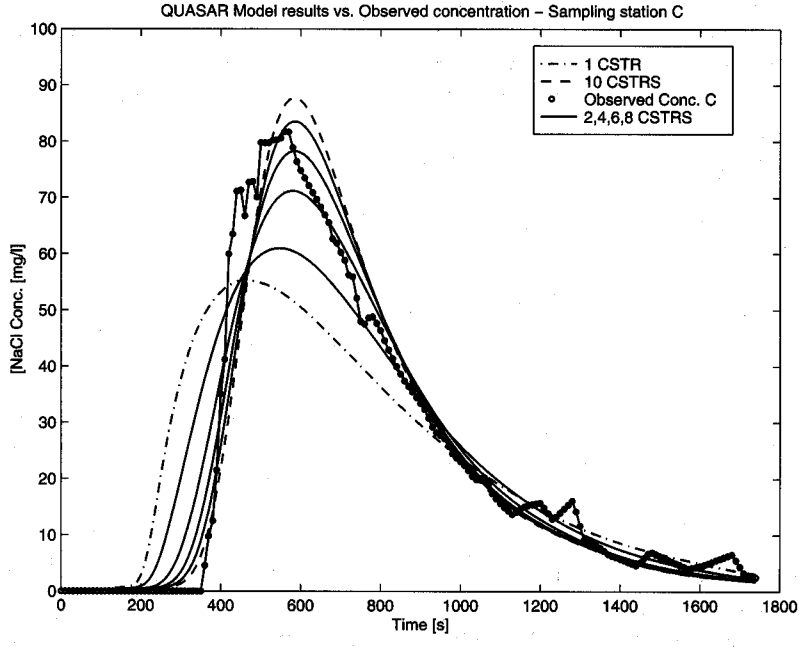

Fig. 8 *Existing QUASAR model results for station C with differing numbers of CSTRs in series*

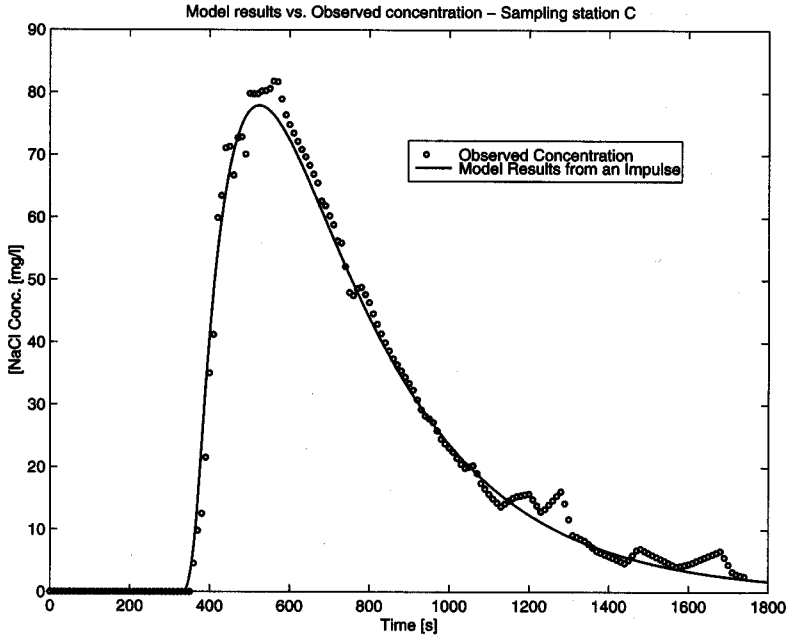

Fig. 9 *Extended QUASAR results with dead-zone parameters estimated from an impulse tracer input using the SRIV method*

It is worth noting, however, that considerable differences result between parameter values obtained from the decomposition and the values identified from the individual analyses performed for reaches B and C (cf. dispersive fraction). MC analysis shows very narrow 95% confidence intervals for the residence time in the three reaches. This result suggests that the serial system decomposition is well identified from the data and therefore the differences in *DF* cannot be explained from the sensitivity of this parameter to model uncertainty. However, note that in the present factorisation example the time delay parameters at the intermediate measurement sites were fixed subjectively. Other values could have been applied, depending on available hydraulic and geometric information, resulting in equally good fits to the downstream observed concentration time curve. This observation is confirmed by performing a sensitivity analysis on the intermediate time delay parameters used. In particular, note that by using the dispersive fraction parameters as identified from the individual analyses for reaches B and C (DF_B = 0.321, DF_C = 0.445) and the residence time parameters identified from the third order factorisation (T_{rA} = 305.53, T_{rB} = 10.25, T_{rC} = 96.03s), a set of time delay parameters (τ_A = 188s, τ_B =22s, τ_C = 120s) are produced that define a serial system with a comparable fit (R^2 = 98%) to the previous analysis shown in Fig. 9.

Conclusions

Incorporation of aggregated dead-zone parameters has resulted in an extended QUASAR model able to characterise effectively and accurately transport mechanisms in non-tidal streams, resulting in a flexible river water quality model. The mathematical representation adopted extends the utility of the existing model to real-time forecasting applications such as the management and control of pollution incidents. Clearly the advantage of the method in forecasting applications results from the incorporation of an explicit lag or time-delay parameter. In this form, the model combines the proven non-conservative pollutant modelling capabilities of QUASAR, with the accurate transport mechanism properties of the ADZ model.

The implemented representation also enables the application of efficient methods of parameter estimation. The SRIV method of system identification provides the means for objective identification of appropriate model reach structures and estimation of associated parameters. In the stream studied here, the estimated time delay and dispersive fraction, together with the identified river structure, accurately describe the observed transport mechanisms. The time delay, the peak and the overall skewed shape of the observed concentration-time curves, were reproduced exceptionally well with the implemented model. The results demonstrate the power of this lumped, data-based mechanistic approach (Young and Lees 1993, Young *et al.*, 1996) when reliable tracer experiment data are available.

To test the applicability of the model to larger rivers, other tests have been performed (Camacho, 1997) with published data from tracer experiments conducted on the Bedford Ouse (Whitehead *et al.* 1986). The extended model typically produced a 91% fit to the downstream tracer curves, which were the result of planned tracer experiments carried out over a range of discharges.

A simple version of this extended QUASAR model has been developed for teaching purposes at the University of

Reading. Contact Whitehead for a copy of this new model HERMES. Work currently in progress at Imperial College is concerned with the integration of data-based mechanistic flow routing and solute transport models. The first two authors would be keen to hear from readers who have any tracer data collected under unsteady flow conditions either in the laboratory or by means of planned field experiments.

Acknowledgements

Luis Camacho is grateful to COLCIENCIAS, the Colombian Institute for Science and Technology Development for financial support granted during 1996/97.

References

Beer, T. and Young, P.C., 1983. Longitudinal dispersion in natural streams, *J.Environ. Engng.*, **109**, No5, ASCE, 1049–1067.

Camacho, L.A., 1997. Extension of the QUASAR water quality model to incorporate dead-zone mixing concepts, *MSc. Dissertation, University of London, Imperial College of Science Technology and Medicine*, London.

Chapra, S.C., 1997. *Surface Water-Quality Modelling*, The McGraw-Hill Companies. Inc., New York.

Doodge, J.C.I., 1973. *Linear theory of hydrologic systems, Agric. Res. Serv.*, USDA, Tech. Bull. 1468, Washington, D.C.

Ferrier R.C., Whitehead, P.G., Sefton, C., Edwards, A.C. and Pugh K., 1995. Modelling impacts of land use change and climate change on nitrate-nitrogen in the River Don, North East Scotland, *Wat. Res.*, **29**, 1950–1956.

Green, H. and Beven, K., 1993. Prediction of times of travel for a pollution incident on the River Eden in March 1993, *CRES Technical Report TR/99*, Lancaster University.

Himmelblau, D.M. and Yates, R.V. 1968. A new method of flood routing, *Wat. Resour. Res.*, 4, No. 6, 1193–1199.

Lewis D.R., Williams, R.J. and Whitehead, P.G., 1997. Quality simulation along river systems (QUASAR): an application to the Yorkshire Ouse, *Sci. Tot. Environ.*, **194/195**, 399–418.

Mathworks, 1996. Matlab/Simulink reference guide, Natick, Mass.

Rutherford, J.C., 1994. *River mixing*, Wiley, Chichester.

Taylor, G.I., 1954. The dispersion of matter in turbulent flow through a pipe, *Proc. Roy. Soc. Lond.*, **223A**, 446–468.

Wallis, S.G., Young, P.C. and Beven, K.J., 1989a. Experimental investigation of the aggregated dead-zone model for longitudinal solute transport in stream channels, *Proc. Inst. Civil Engrs.*, Part 2, **87**, 1–22.

Wallis, S.G., Guymer, I. and Bilgi, A., 1989b. A practical engineering approach to modelling longitudinal dispersion, *Proc. of Int. Conf. on hydraulic and environmental modelling of coastal, estuarine and river waters*, Bradford, England 1–21 Sept., 291–300.

Wallis, S.G., 1994. Simulation of solute transport in open channel flow. In Mixing and transport in the environment (Ed. K. Beven, P. Chatwin, J. Millbank). Wiley, Chichester, pp. 89–112.

Weinmann, P.E. and Laurenson, E.M., 1979. Approximate flood routing methods: A review, *J. Hydraul. Div.*, **105**, No12, ASCE, 1521–1536.

Whitehead, P.G., Young, P. and Hornberge,r G., 1979. A systems model of stream flow and water quality in the Bedford-Ouse River-1. Stream flow modelling, *Wat. Res.*, **13**, 1155–1169.

Whitehead, P.G., Beck, B., and O'Connell, P.E., 1981. A systems model of stream flow and water quality in the Bedford-Ouse River-II. Water quality modelling, *Wat. Res.*, **15**, 1157–1171.

Whitehead, P.G., and Williams R.J., 1982. A dynamic nitrogen balance model for river systems. *IAHS Conf. Proc*, IAHS publication 139.

Whitehead, P.G., Caddy D.E. and Templeman R.F., 1984. An on-line monitoring, data management and forecasting system for the Bedford Ouse River Basin. *Wat. Sci. Technol.*, **16**, 295–314.

Whitehead, P.G., Williams, R.G., and Hornberger, G.M., 1986. On the identification of pollutant or tracer sources using dispersion theory, *J. Hydrol.*, **84**, 273–286.

Whitehead, P.G., McCartney, M.P., Williams, R.J., Ishemo C.A.L. and Thomas R. 1995. A method to simulate the impact of acid mine drainage on river systems, *J. Chartered Inst. Wat. Engrs. Managers*, **9**, 119–131.

Whitehead, P.G., Williams, R.J. and Lewis, D.R. 1997. Quality simulation along river systems (QUASAR): model theory and development, *Sci. Tot. Environ.*, **194/195**, 447–456.

Young, P.C., 1984. *Recursive Estimation and Time-Series Analysis*, Springer-Verlag, Berlin.

Young P.C., 1991. Simplified Refined Instrumental Variable (SRIV) estimation and True Digital Control (TDC): a tutorial introduction, *Proc. First European Control Conference*, Grenoble, 1295–1306.

Young, P.C., 1992. Parallel processes in hydrology and water quality: a unified time-series approach, *J. Inst. Wat. Engrs. Managers*, **6**, 598–612.

Young, P.C. and Wallis, S.G., 1993. Solute Transport and Dispersion in Channels. In *Channel Network Hydrology* (Eds. K. Beven and M.J. Kirby). Wiley, Chichester, pp. 129–174.

Young, P.C. and Lees, M.J., 1993. The active mixing volume, a new concept in modelling environmental systems. In *Statistics for the Environment.* (Ed. V. Barnett and K. Turkman). Wiley, Chichester, pp. 3–43.

Young, P.C., Parkinson, S. and Lees, M.J., 1996. Simplicity out of complexity in environmental modelling: Occam's razor revisited. *J. App. Stat.*, **23**, 165–210.